

Semantic Web Technologies I

Lehrveranstaltung im WS12/13

PD Dr. Sebastian Rudolph
Dr. Duc Thanh Tran
M.Sc. Anees ul Mehdi

Linked Data

Dr. Duc Thanh Tran

XML und URIs

Einleitung in RDF

RDF Schema

Logik – Grundlagen

Semantik von RDF(S)

SPARQL – Syntax und Intuition

Semantik von SPARQL

Linked Data

Semantic Search

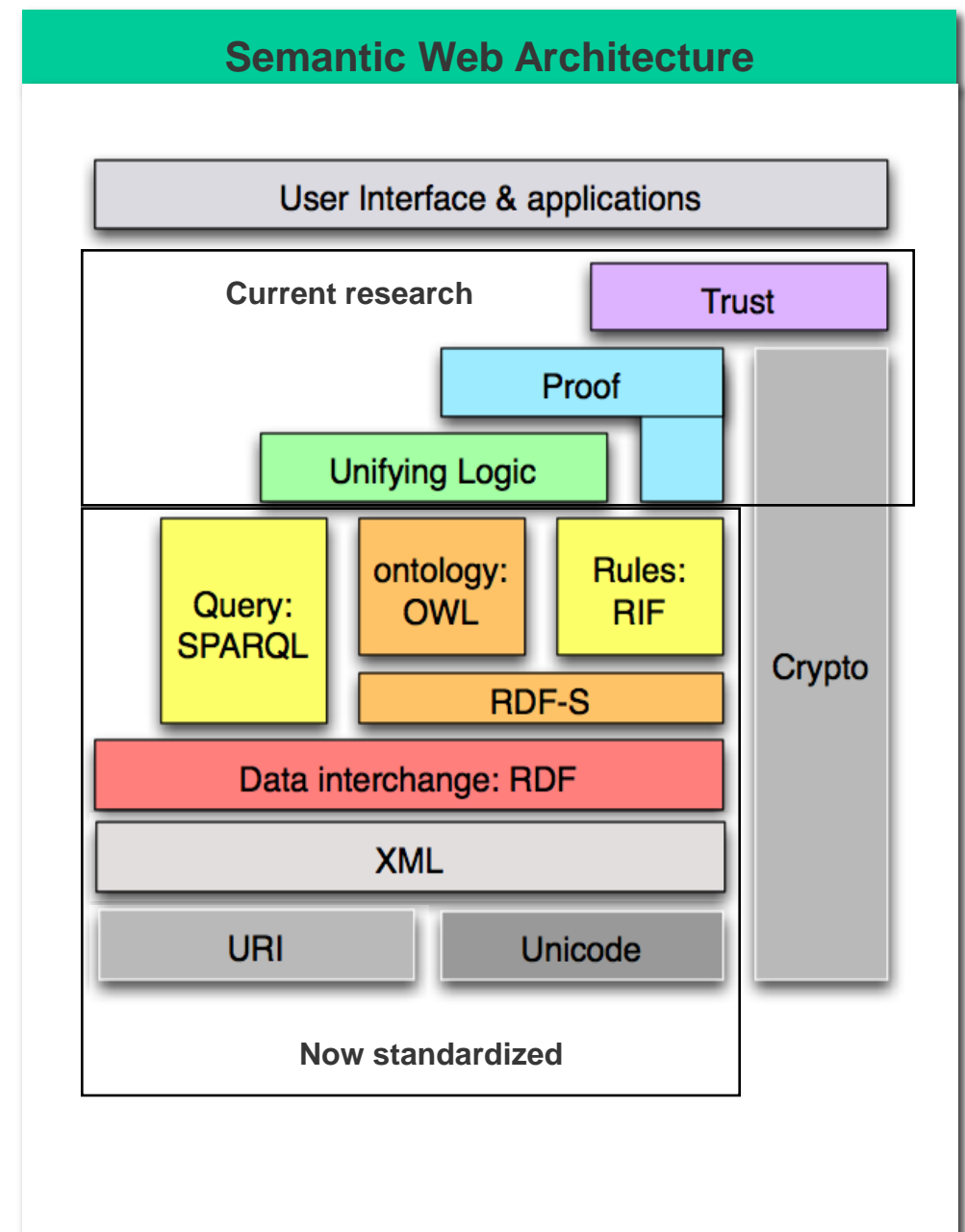
OWL – Syntax und Intuition I

OWL – Syntax und Intuition II

OWL – Semantik und Reasoning

Konjunktive Anfragen und Regelsprachen

Applications



Agenda

- Motivation
- Linked Data Principles
- Linked Data Access

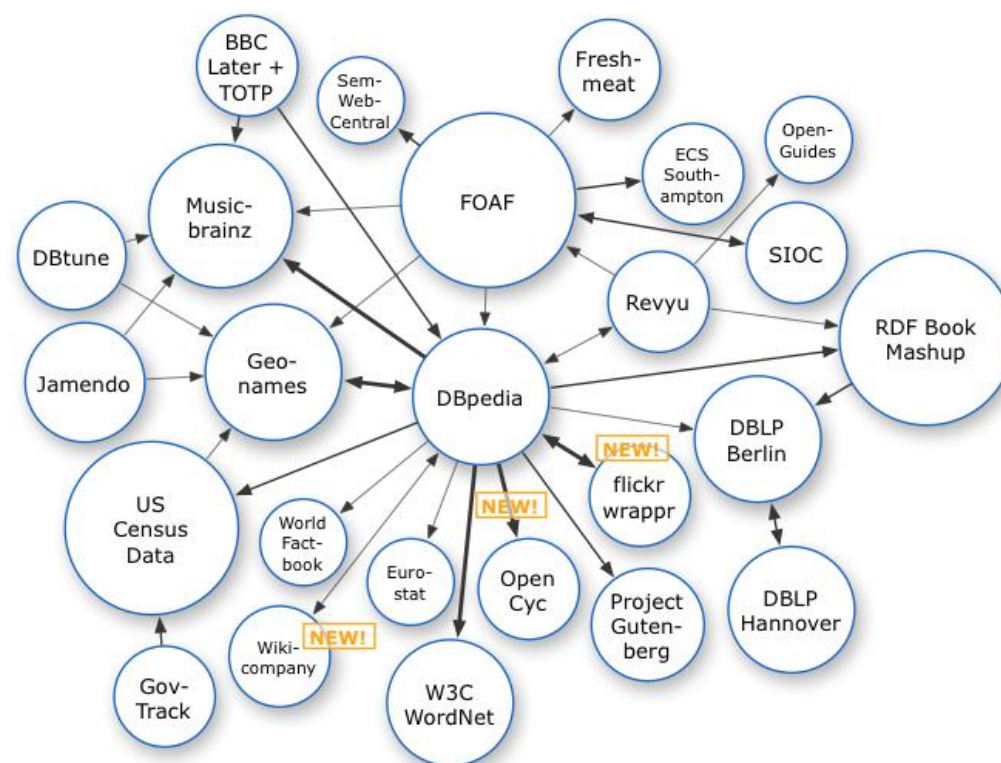
MOTIVATION

Motivation

AIFB 

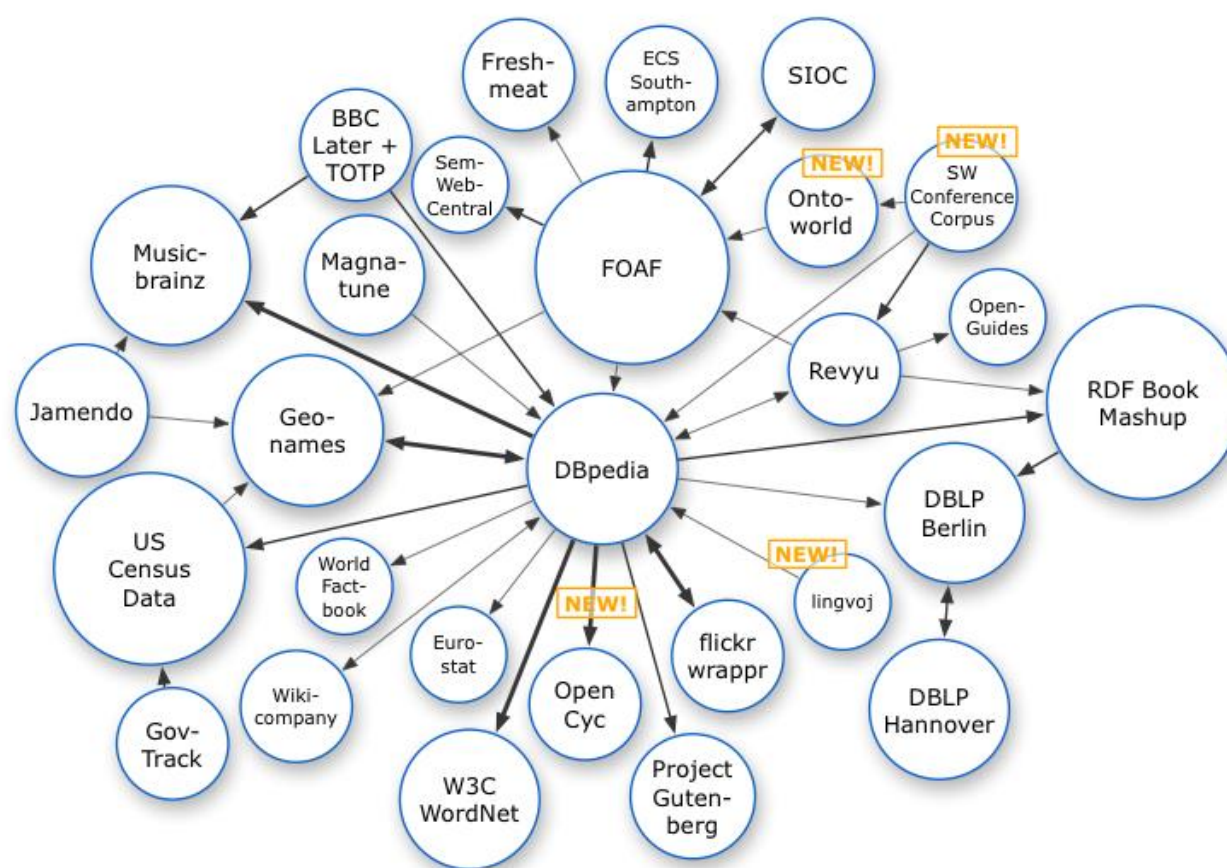
- Data volumes explode
 - More and more **data available on the Web is represented in Semantic Web standards**
 - Linking Open Data (LOD) initiative
- Semantic Web standards & technologies facilitate
 - Representation, exchange
 - Integration & linking of data from multiple sources
 - Consumption

Linked Data on the Web



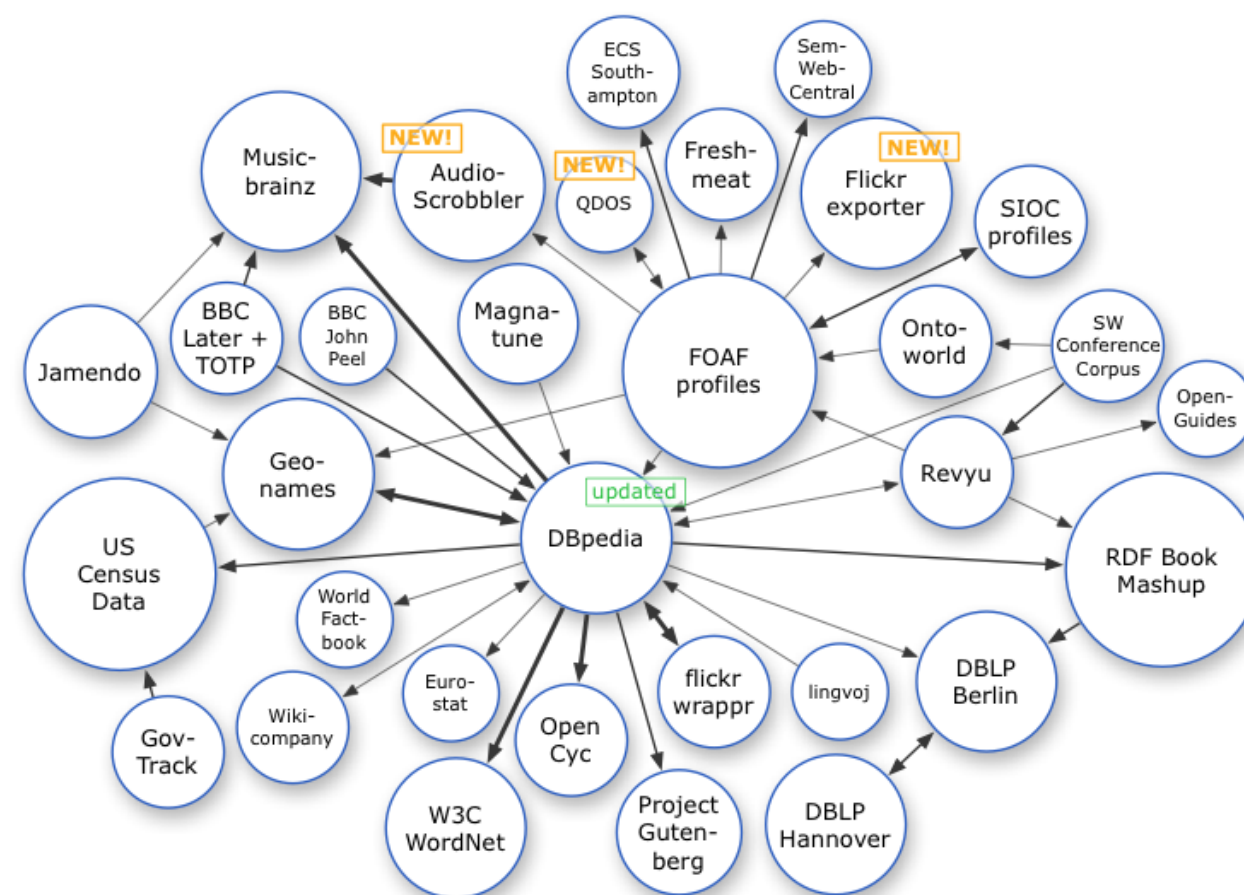
2007-10

Linked Data on the Web



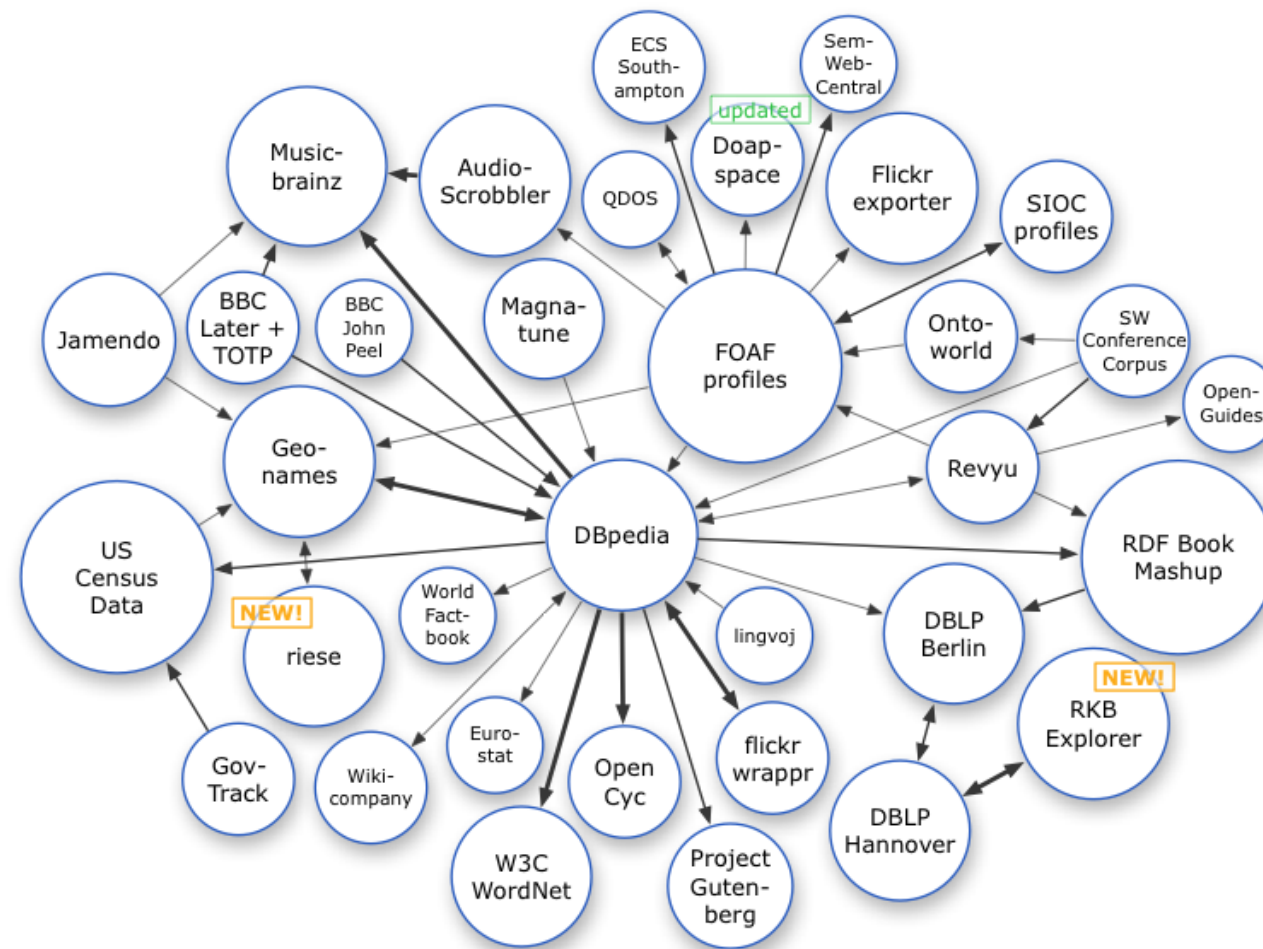
2007-11

Linked Data on the Web



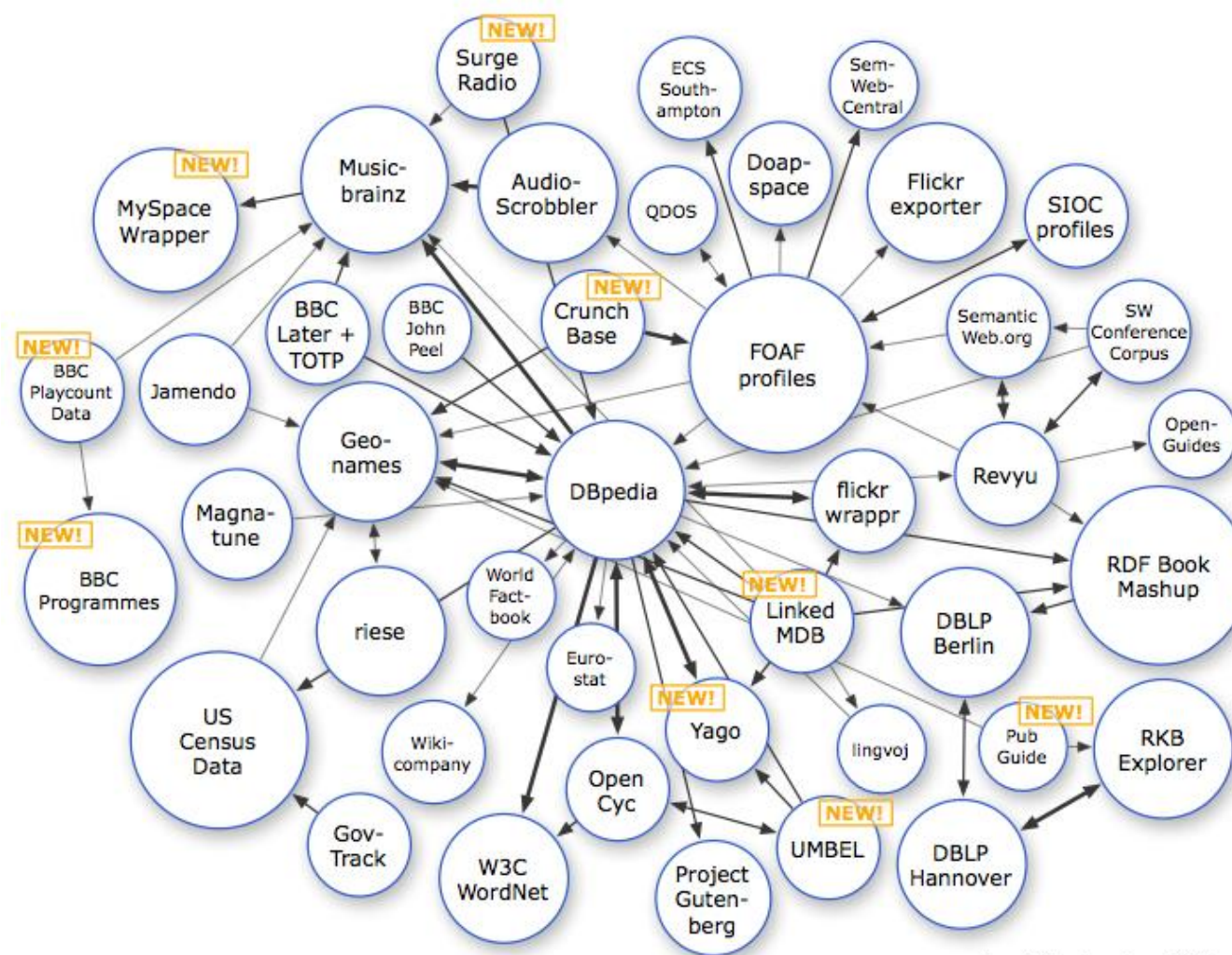
2008-02

Linked Data on the Web



2008-03

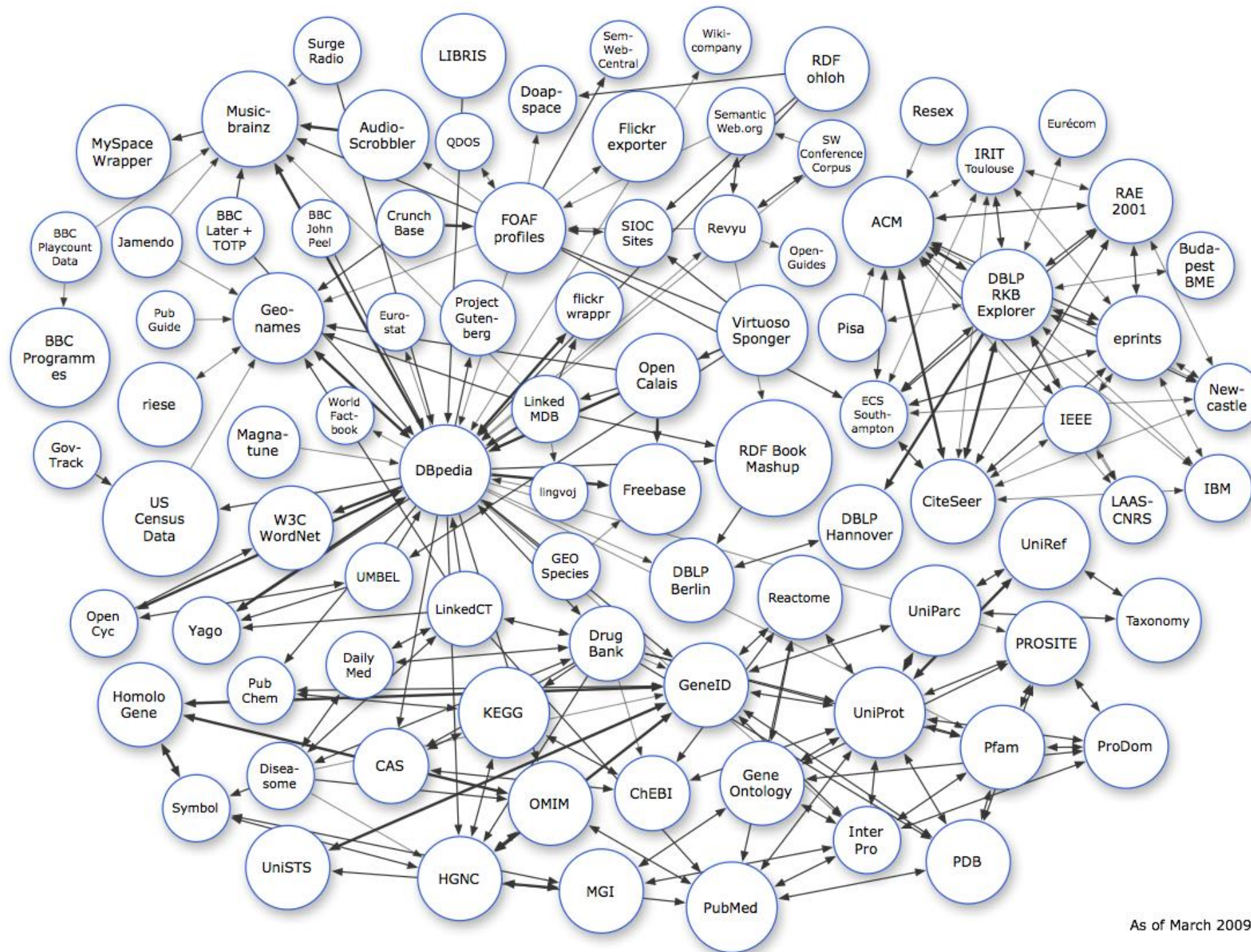
Linked Data on the Web



As of September 2008

2008-09

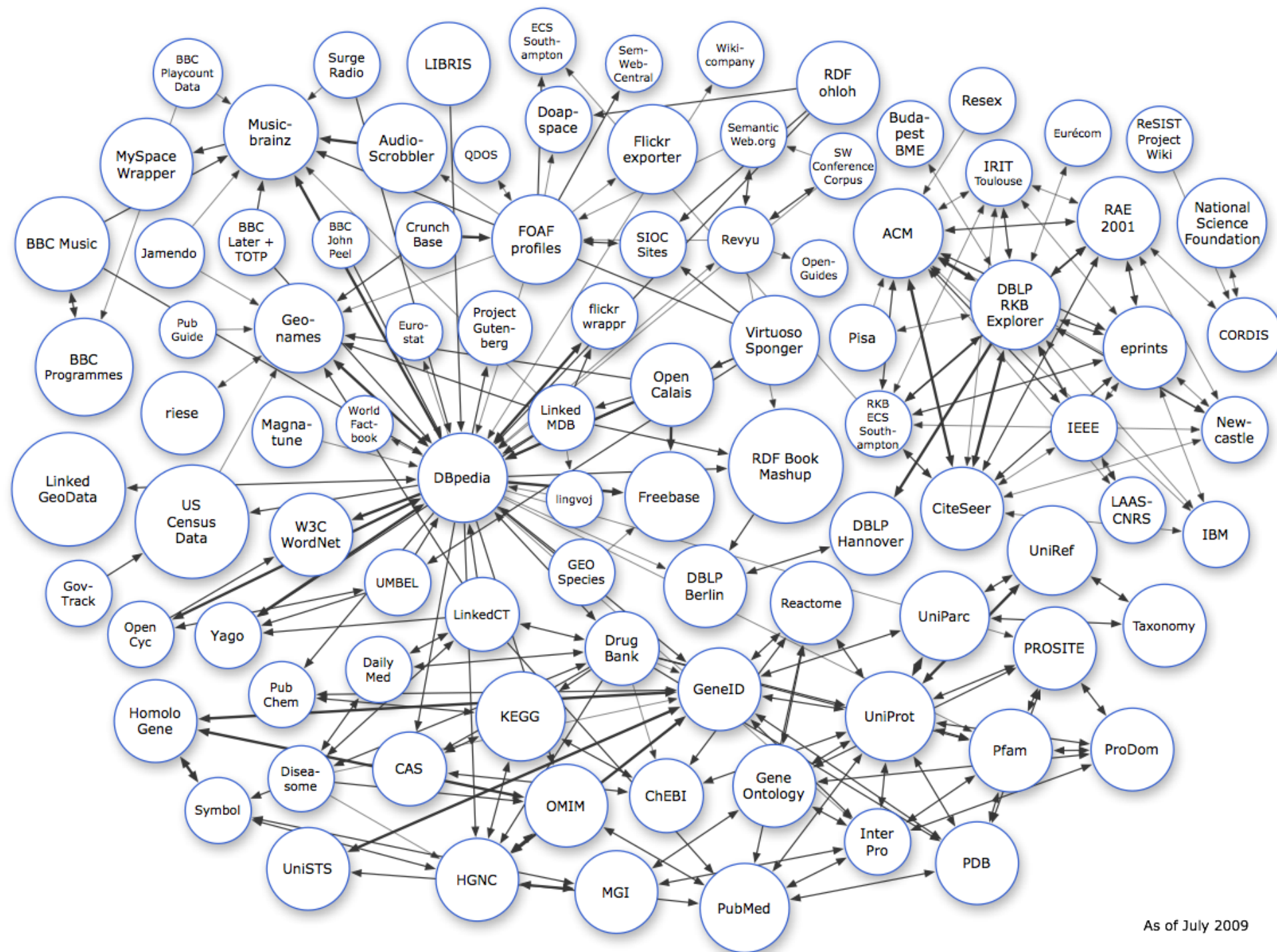
Linked Data on the Web



As of March 2009

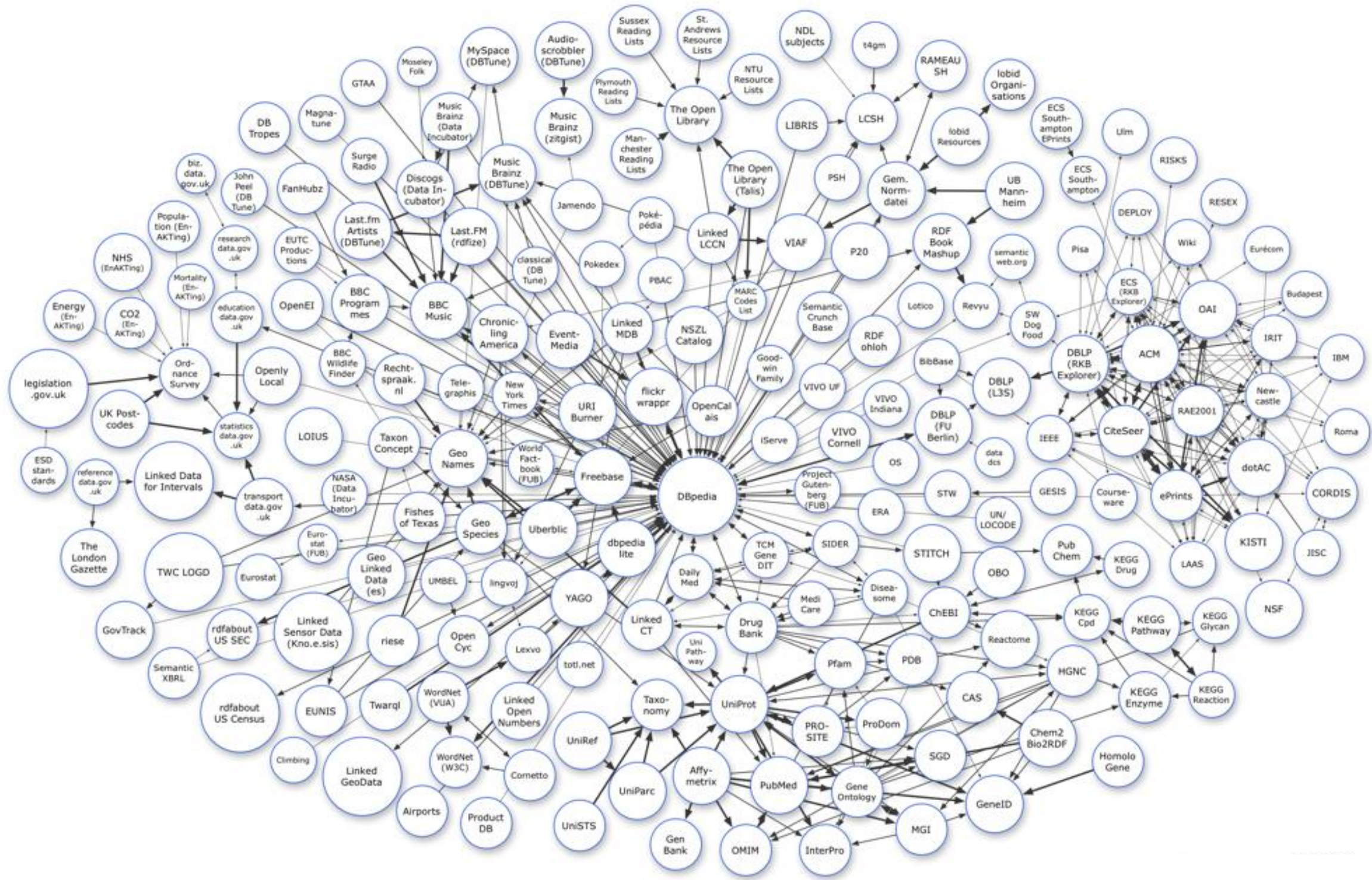
2009-03

Linked Data on the Web



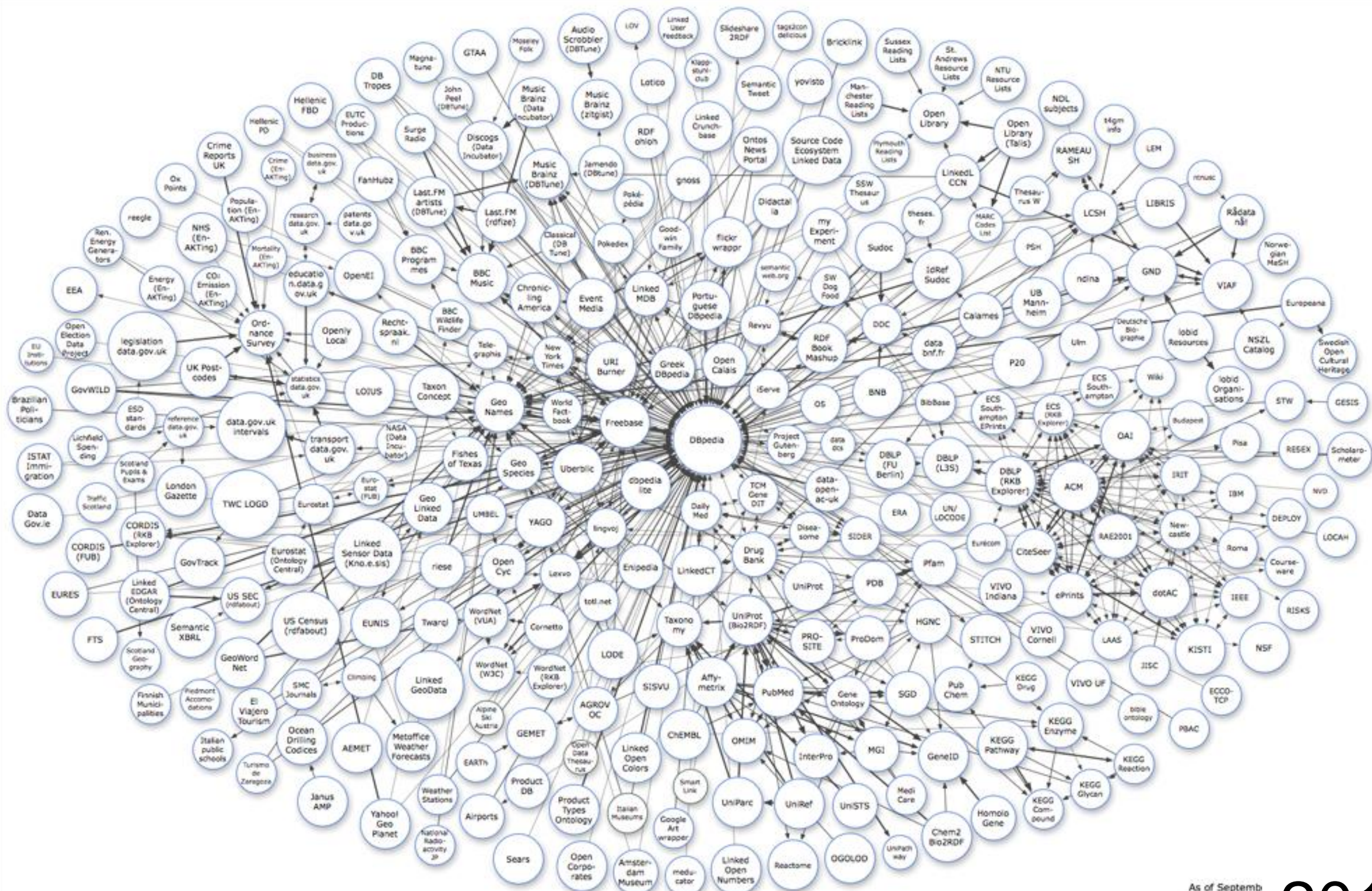
2009-07

Linked Data on the Web



2010-09

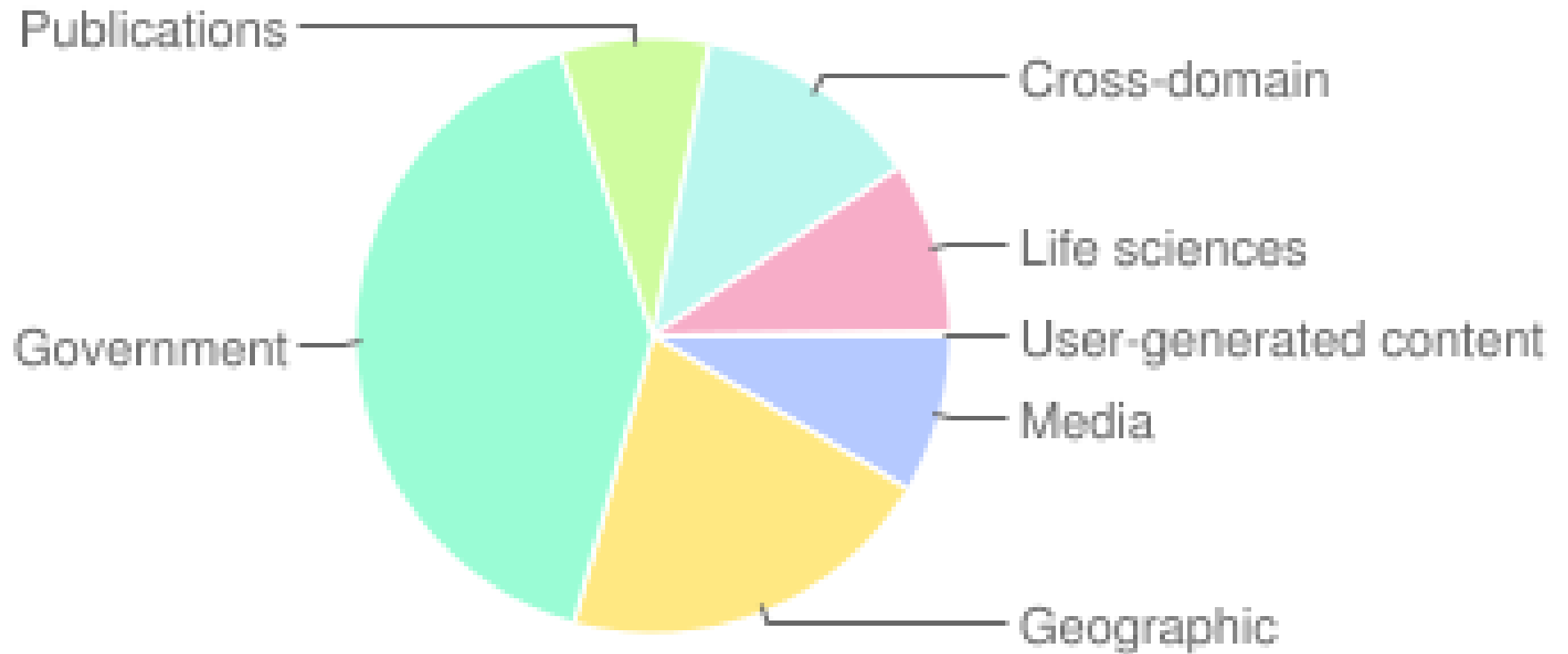
Linked Data on the Web



As of Septemb

2011-09

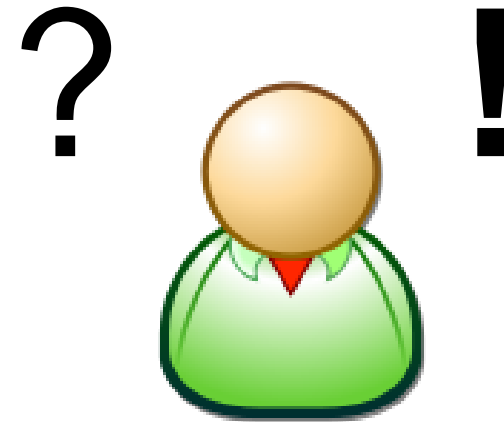
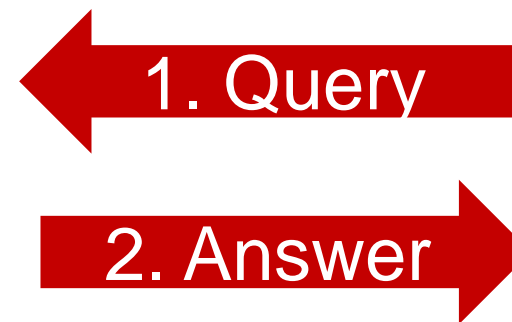
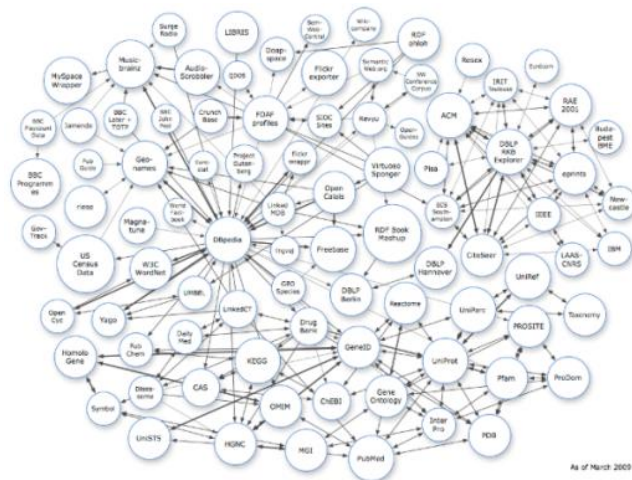
Types of Data in the Linking Open Data Cloud



<http://www4.wiwiiss.fu-berlin.de/lodcloud/state/> (Sept 2010)

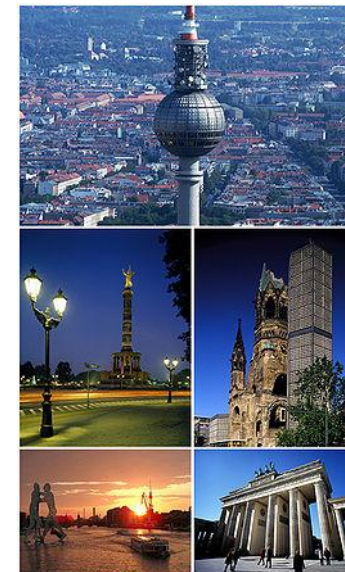
Consumption

A Search Scenario



Semantic Technologies facilitate access to data

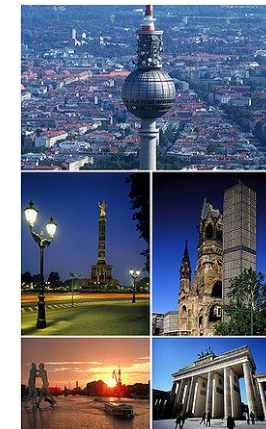
- Q: data about Berlin?
- Q: famous people that died in Berlin?
- Q: data about Hegel?
- Q: Hegel's publications?
- Q: data about Marlene Dietrich?
- Q: Dietrich's songs?



DBpedia



- Linked Data version of Wikipedia
- Scripts that extract data (text, links, infoboxes) from Wikipedia
- Published as Linked Data
- Interlinking hub in the Linked Data web
- Berlin
 - <http://dbpedia.org/resource/Berlin>
- Hegel
 - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
- Marlene Dietrich
 - http://dbpedia.org/resource/Marlene_Dietrich



BBC Music

AIFB 

- Data about BBC (radio) programmes, artists, songs...
- Combination of BBC-internal data (playlists), MusicBrainz (artists, albums), Wikipedia (artists)
- Underpinning the BBC Music website
- Data published according to Linked Data principles
- Marlene Dietrich
 - <http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5>



Virtual International Authority File (VIAF)



- Joint project of national libraries and related organisations
 - 21 institutions, among them the Deutsche Nationalbibliothek
- Provide access to “authority files”
- Matching and interlinking collections from participating institutions
- Hegel
 - <http://viaf.org/viaf/89774942>
- Marlene Dietrich
 - <http://viaf.org/viaf/97773925>

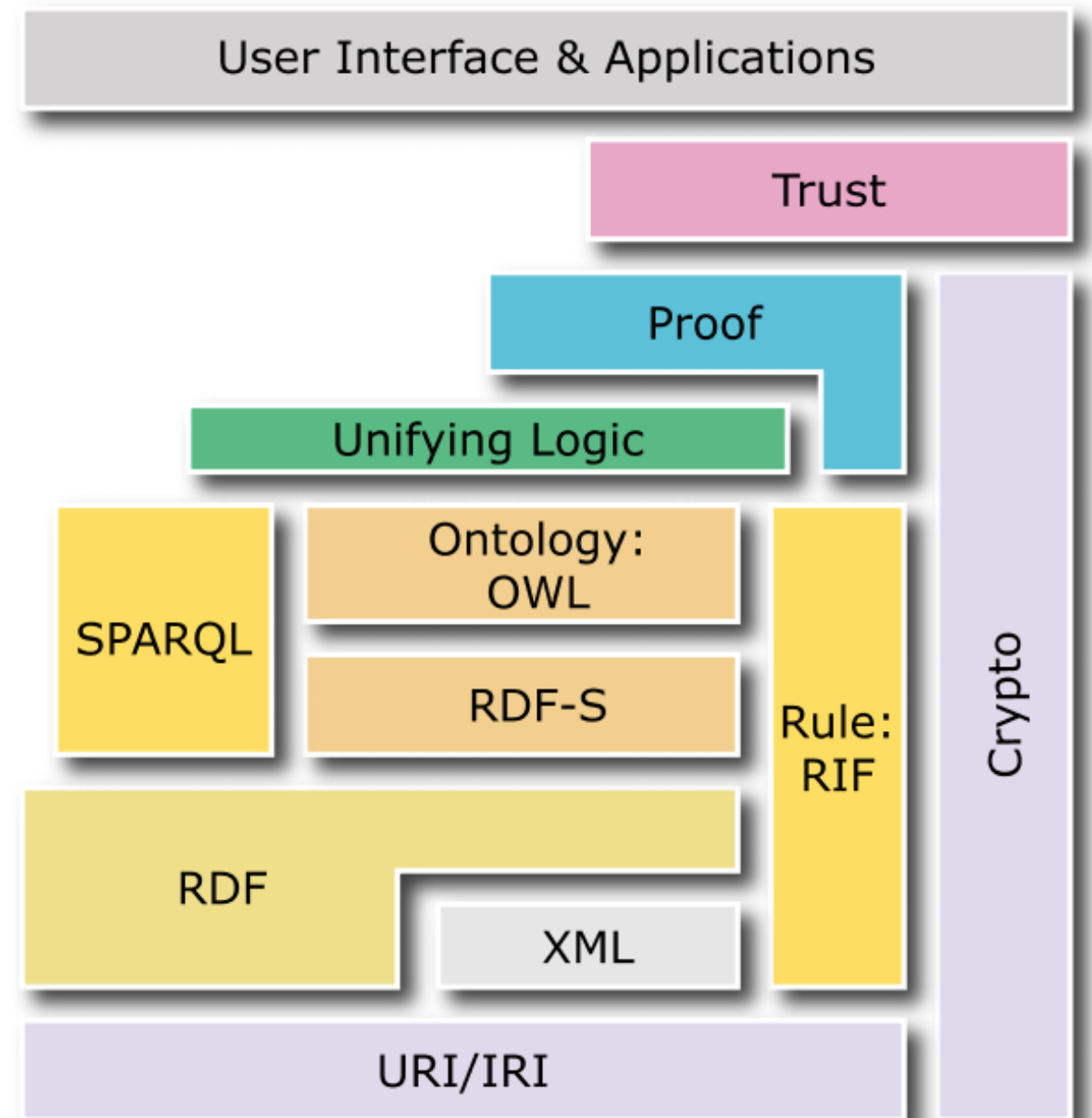


LINKED DATA PRINCIPLES

Semantic Technologies



- Semantic Web technologies, standardised by the W3C, are mature:
 - RDF recommendation in 1999, update in 2004
 - RDFa (RDF in HTML) note in 2008
 - RDFS recommendation in 2004
 - SPARQL recommendation in 2008
 - OWL recommendation in 2004, update in 2009
- Linked Data is a subset of the Semantic Web stack, including web architecture:
 - HTTP
 - IRI / URI
 - (RDF / XML)



Linked Data Principles



1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

<http://www.w3.org/DesignIssues/LinkedData>

1. Use URIs as Names for Things

- Use a unique identifier (URIs) to denote things
- Hegel, Georg Wilhelm Friedrich
 - http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
 - <http://viaf.org/viaf/89774942>
 - ...



Names for Things



*“Now! That should clear up
a few things around here!”*

2. Use HTTP URIs

AIFB 

- Enables “lookup” of URIs
- Via Hypertext Transfer Protocol (HTTP)
- Piggy-backs on hierarchical Domain Name System to guarantee uniqueness of identifiers
- Uses established HTTP infrastructure
- Connects logical level (thing) with physical level (source)
- Important: distinction between name/“thing URI” and location/“source URI” („other resource“/„non-information resource“ vs. „information resource“)

Information Resources vs. Other Resources

AIFB 

Marlene Dietrich, the person



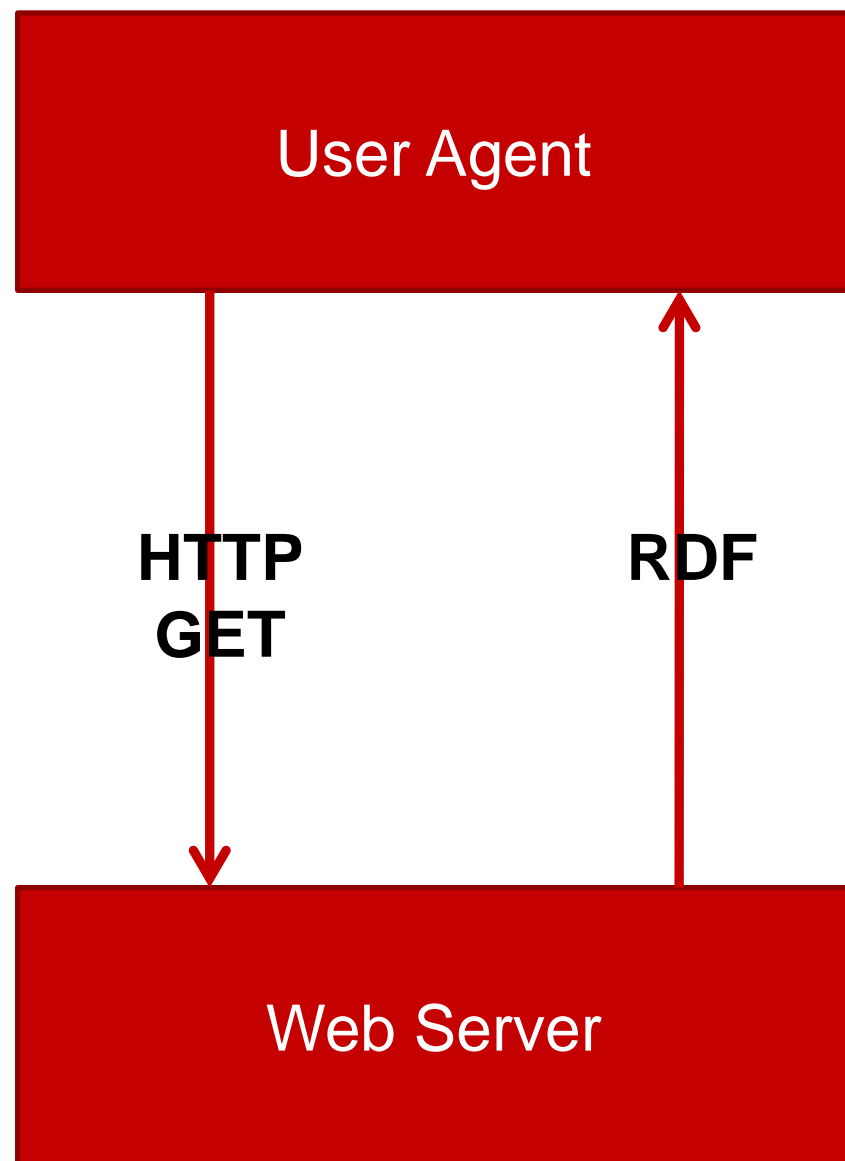
Name?
Creator?
Birth date?
Last change date?
License?
Copyright?

...

File containing data about
Marlene Dietrich



Correspondence between thing-URI and source-URI



<http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5#artist>



<http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5.rdf>

3. Provide Useful Information

- When somebody looks up a URI, return data using the standards (RDF*, SPARQL)
- Resource Description Framework, a format for encoding graph-structured data (with URIs to identify nodes/vertices and links/edges)

Resource Description Framework

- Directed, labeled graph
- `triple(subject, predicate, object)`
 - subject: URI (or blank node)
 - predicate: URI
 - object: URI (or blank node) or RDF literal (string, integer, date...)
- RDF/XML is the most widely deployed serialisation
- Other serialisations possible (N-Triples, Turtle, Notation3...)

4. Link to Other URIs



- Enable people (and machines) to jump from server to server
- External links vs. internal links (for any predicate)
- Using external vocabularies enables linking
- Vocabularies might be interlinked, too
- Special owl:sameAs links to denote equivalence of identifiers (useful for data merging)

Equivalences via owl:sameAs



<http://viaf.org/viaf/89774942>

- http://dbpedia.org/resource/Georg_Wilhelm_Friedrich_Hegel
- <http://www.idref.fr/026917467/id>
- <http://libris.kb.se/resource/auth/190350>
- <http://d-nb.info/gnd/118547739>



<http://www.bbc.co.uk/music/artists/191cba6a-b83f-49ca-883c-02b20c7a9dd5#artist>

- http://dbpedia.org/resource/Marlene_Dietrich

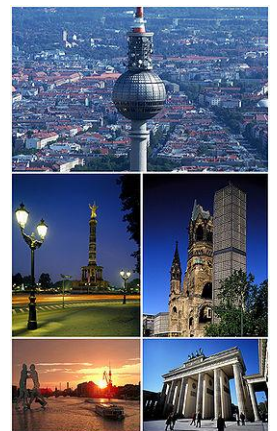
<http://viaf.org/viaf/97773925>

- http://dbpedia.org/resource/Marlene_Dietrich
- <http://d-nb.info/gnd/118525565>
- <http://libris.kb.se/resource/auth/238817>
- <http://www.idref.fr/027561844/id>



<http://dbpedia.org/resource/Berlin>

- <http://mpii.de/yago/resource/Berlin>
- <http://data.nytimes.com/N50987186835223032381> - Berlin (Germany)
- <http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/Berlin>
- <http://data.nytimes.com/16057429728088573361> - Gaspé Peninsula (Quebec) (?)



Benefits of Linked Data



- Explicit, simple data representation
 - Common data representation (Resource Description Framework, RDF) hides underlying technologies and systems
- Distributed System
 - Decentralised distributed ownership and control facilitates adoption and scalability
- Loose coupling with common language layer
 - Large scale systems require loose coupling, via HTTP as common access protocol
- Ease of publishing and consumption
 - Simple and easy-to-use systems and technologies to facilitate uptake
- Cross-referencing
 - Allows for linking and referencing of existing data, via reuse of URIs
- Incremental data integration
 - Start with merged RDF graphs and provide mappings as you go

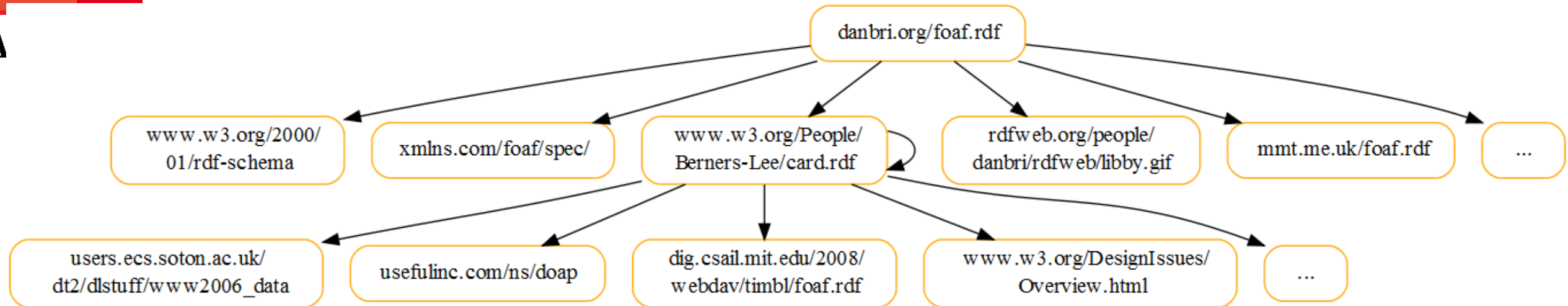
Challenges

AIFB 

- Ramp-up cost for data conversion
 - May be alleviated by semi-automatic mappings and adequate tool support for manual conversion
- Integrated data may be messy at first
 - But can be refined as need arises
- Distributed creation and loose coordination may result in inconsistencies
 - Can be detected, diagnosed, and fixed with appropriate tools
- Not only quality but also volume is a problem

Resource Graph Explosion

A



Directed graph rooted in `http://danbri.org/foaf.rdf`

Level 0: 1

Level 1: 25^1 (avg), 105^1 (worst)

Level 2: 25^2 (avg), 105^2 (worst)

Level 3: 25^3 (avg), 105^3 (worst)

Level k: n^k

ACCESSING LINKED DATA

APPLICATIONS

CRAWLING

INDEXING

QUERY PROCESSING

Basic Application: Entity Browsing



Warehousing/ Crawl-Index-Serve

SWSE

India New Search

[RDF](#) [next ▶](#) Results 1 - 1

India
<http://dbpedia.org/resource/India>

<p>TITLE</p> <ul style="list-style-type: none"> India Intia, Indie, Inde, Índia, इन्द, Индия, Indien, Republic of India, India Intia, Indie, Índia, Inde, इन्द, Hindistan, Индия, Indien, Republic of India, 印度, Индія, India <p>LATITUDE</p> <ul style="list-style-type: none"> 21.0^^http://www.w3.org/2001/XMLSchema#double 28.613333^^http://www.w3.org/2001/XMLSchema#float 20.0 <p>LONGITUDE</p> <ul style="list-style-type: none"> 77.208336^^http://www.w3.org/2001/XMLSchema#float 78.0^^http://www.w3.org/2001/XMLSchema#double 	<p>IS IN SCHEME</p> <ul style="list-style-type: none"> nytd_geo <p>SAMEAS</p> <ul style="list-style-type: none"> Mx4rvVijnZwpEbGdrcN5Y29ycA de588909794eba0786311e211f0e466dab: d095589843f71ea26ceb90be57071834bbe <p>TYPE</p> <ul style="list-style-type: none"> SpatialThing Thing 1 2094 Country <p>SUBJECT</p> <ul style="list-style-type: none"> Countries of the Indian Ocean English-speaking countries and territories Federal countries <p>CHILDREN FEATURES</p>
--	---

Virtual Integration/ Distributed Querying

URI: <http://dbpedia.org/resource/India> [Help About](#) Sources pending: -26

This is release 0.8 of the Tabulator Project. The [live development trunk](#) is also available.

▼ India - Wikipedia, the free encyclopedia

request

- Request for <http://dbpedia.org/resource/India>
- Request for <http://mpii.de/yago/resource/India>
- Request for <http://www.mpi-inf.mpg.de/yago/resource/India>

requested by

- India - Wikipedia, the free encyclopedia

title

- India - Wikipedia, the free encyclopedia

seeAlso

- /w/index.php?title=Special:RecentChanges&feed=atom
- http://dbpedia.org/data/India.xml
- http://yago.zitgist.com/India

sameAs

- India - Wikipedia, the free encyclopedia
- India - Wikipedia, the free encyclopedia

is assembly of

- http://dbpedia.org/resource/Audi_A4_(B8)
- http://dbpedia.org/resource/Audi_Q5
- http://dbpedia.org/resource/BMW_3_Series
- http://dbpedia.org/resource/BMW_3_Series_(E90)
- http://dbpedia.org/resource/Chevrolet_Tavera
- http://dbpedia.org/resource/Fiat_Grande_Punto
- http://dbpedia.org/resource/Fiat_Palio
- http://dbpedia.org/resource/Fiat_Siena
- http://dbpedia.org/resource/Ford_Escort_(Europe)_Ford_Escort_Mark_VI
- http://dbpedia.org/resource/Ford_Everest

32 more...

SWSE, Falcons, Sindice, Watson,
FactForge...

Tabulator, Disco, Zitgist...

Data Sources in SemanticSearch@AIFB Demo

[SWC Finalist 2010]



- English Wikipedia
- Data from Linked Open Data
 - DBpedia
 - YAGO
 - Many more
- Live data from Data.gov (US Government)
 - E.g. live data about earthquakes



Wikipedia + DBpedia

Alice (album)

View Blog Revisions

Alice is an album by [Tom Waits](#), released in 2002 on [Epitaph Records](#) (under the *Anti* sub-label). The album contains the majority of songs written for the play *Alice*, based on the forbidden love between [Lewis Carroll](#) and [Alice Liddell](#), for whom he wrote the story *Alice's Adventures in Wonderland*. The adaptation was directed by [Robert Wilson](#), whom Waits had previously worked with on the play *The Black Rider*, and originally set up at the [Thalia Theatre](#) in [Hamburg](#) in 1992. The play has since been performed in various theatres around the world.

The album was co-released with *Blood Money*, containing songs from a play adapted by Robert Wilson from [Georg Büchner's](#) *Woyzeck*.

It was ranked #2 in [Metacritic's](#) Top 30 albums of 2002.^[1]

The songs had been released as a [bootleg](#) in several different versions called *The Alice Demos* many years before its official release. The source is believed to be studio recordings taken when Waits' car was broken into in late 1992.^[2]

Contents

[Chart information](#)
[Track listing](#)
[Musicians](#)
[References](#)

Chart information

Chart	Peak position
Switzerland	24
UK	20
US	33

Track listing

All tracks written by Tom Waits and [Kathleen Brennan](#).

- "Alice" – 4:28
 - Tom Waits - Vocal, Piano
 - Eric Perney - Bass
 - Colin Stetson - Sax

Image

YouTube

Tom Waits - Kommienezuspadt

source: <http://iwb.fluidops.com>

Wikipedia + DBpedia

Alice (album)

Predicate	Value	In
release_date	2002-05-04	
subsequent work	Real Gone	
depiction	Tom Waits-Alice.jpg	
skosCore04:subject	2002 albums	
skosCore04:subject	Tom Waits albums	
skosCore04:subject	Soundtrack albums	
skosCore04:subject	Steampunk music	
skosCore04:subject	Music based on Alice in Wonderland	
skosCore04:subject	ANTI- albums	
artist	Tom Waits	
genre	Rock and roll	
name	Alice	
label	Alice (album)	
runtime (s)	2880.0	
record label	ANTI-	
record label	Epitaph Records	
type	Album	
review	http://www.robertchristgau.com/get_artist.php?name=Tom Waits	
review	http://www.allmusic.com/cg/amg.dll?p=amg&sql=A2hq2g4fptv3z	
review	http://avclub.com/content/node/13355	
review	http://pitchforkmedia.com/article/record_review/23067/Tom_Waits_Alice_and_Blood_Money	
review	http://www.musicemissions.com/artists/albums/index.php?album_id=782	
review	http://www.rollingstone.com/artists/tomwaits/albums/album/114401/review/5943251/alice	
type	musical work	
type	Thing	

Image

YouTube

Tom Waits - Kommienezuspadt

source: <http://iwb.fluidops.com>

Query Interpretation, Refinement and Exploration

Search Results

Keywords

queen single

Search

Click on one of the suggestions to initiate translation! (can take a few seconds)

queen single

Set searchfield to "queen single"

A (queen) is a Single

B writer A (queen)

B is a Single

A is a Single

A producer B (queen)

Semantic Completions

queen single.php

queen singled

queen singler

queen singlerpt

queen singles

queen singles-1997-2007

queen singles/2002/03/04/the

queen singles/2002/03/25/new

Queen (band)

Queen (band)

Queen (band)

Queen (band)

Queen (band)

Queen (band)

Queen (band)

Syntactic Completions

RESULT COLUMN1

producer

Range: All Values (43)

type

Range: All Values (43)

writer

Range: All Values (42)

Musical Artist (42)

Brian May (13)

Frank Musker (1)

Freddie Mercury (14)

John Deacon (7)

Roger Meddows-Taylor (7)

Facets

Initial Query

See Entire Query

?sx1

A Kind of Magic (song)

Another One Bites the Dust

Back Chat

Bicycle Race

Body Language (song)

Calling All Girls

Crazy Little Thing Called Love

Fat Bottomed Girls

Good Old-Fashioned Lover Boy

Hammer to Fall

Heaven for Everyone

I Want to Break Free

It's Late

It's a Hard Life

Keep Yourself Alive

Killer Queen

Las Palabras de Amor

Liar (Queen song)

Long Away

Mustapha

Result Inspection, Analysis and Browsing

Earthquake

Login / Register

Search

fluid Operations

IWB Tabs

Semantic Wiki Table Graph

View Blog Edit Revisions

An **earthquake** (also known as a **tremor** or **temblor**) is the result of a sudden release of energy in the **Earth's crust** that creates **seismic waves**. Earthquakes are recorded with a **seismometer**, also known as a seismograph. The **moment magnitude** of an earthquake is conventionally reported, or the related and mostly obsolete **Richter magnitude**, with magnitude 3 or lower earthquakes being mostly **imperceptible** and magnitude 7 causing serious damage over large areas. Intensity of shaking is measured on the modified **Mercalli scale**.

At the Earth's surface, earthquakes manifest themselves by shaking and sometimes displacing the ground. When a large earthquake **epicenter** is located offshore, the seabed sometimes suffers sufficient displacement to cause a **tsunami**. The shaking in earthquakes can also trigger landslides and occasionally volcanic activity.

In its most generic sense, the word *earthquake* is used to describe any seismic event — whether a natural **phenomenon** or an event caused by humans — that generates seismic waves. Earthquakes are caused mostly by rupture of geological **faults**, but also by volcanic activity, landslides, mine blasts, and nuclear experiments. An earthquake's point of initial rupture is called its **focus** or **hypocenter**. The term **epicenter** refers to the point at ground level directly above the hypocenter.

Contents

- Naturally occurring earthquakes
 - Earthquake fault types
 - Earthquakes away from plate boundaries
 - Shallow-focus and deep-focus earthquakes
 - Earthquakes and volcanic activity
 - Earthquake clusters
 - Aftershocks
 - Earthquake swarms
 - Earthquake storms
- Size and frequency of occurrence
- Induced seismicity
- How to measure and locate an earthquake
- Effects/impacts of earthquakes
 - Shaking and ground rupture
 - Landslides and avalanches
 - Fires
 - Soil liquefaction
 - Tsunami
 - Floods
 - Human impacts
- Preparation
- History
 - Pre-Middle Ages

Preliminary Determination of Epicenters
358,214 Events, 1963 - 1998

Global earthquake **epicenters**, 1963–1998

Query Results

Input

Output

Aggregation

datetime

magnitude

None

Query Results

Sunday, October 25, 2009 04:09:42 UTC	4.5
Sunday, October 25, 2009 03:30:41 UTC	4.2
Sunday, October 25, 2009 03:04:48 UTC	3.8
Sunday, October 25, 2009 03:04:22 UTC	4.1
Sunday, October 25, 2009 02:54:49 UTC	4.0
Sunday, October 25, 2009 02:44:22 UTC	4.1
Sunday, October 25, 2009 02:41:49 UTC	4.0
Sunday, October 25, 2009 02:31:57 UTC	3.5
Sunday, October 25, 2009 01:50:44 UTC	4.2
Sunday, October 25, 2009 01:50:01 UTC	4.8

GMap Earthquake

Karte

Satellit

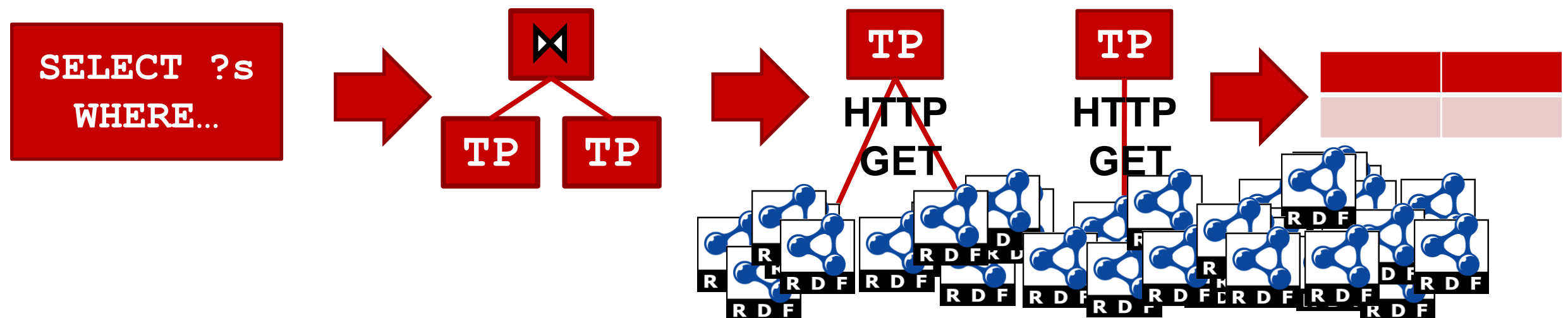
Hybrid

Querying Data Across Sources

- Data warehousing or materialisation-based approaches (MAT)



- Distributed query processing approaches (DQP) over Linked Data sources



- DQP over RDF stores (not covered)

(Linked Data) Crawling

AIFB 

1. Get URI from a queue
2. Open connection and fetch content
3. Process and store content
4. Extract new links and put into queue
5. At defined intervals: schedule URIs in queue

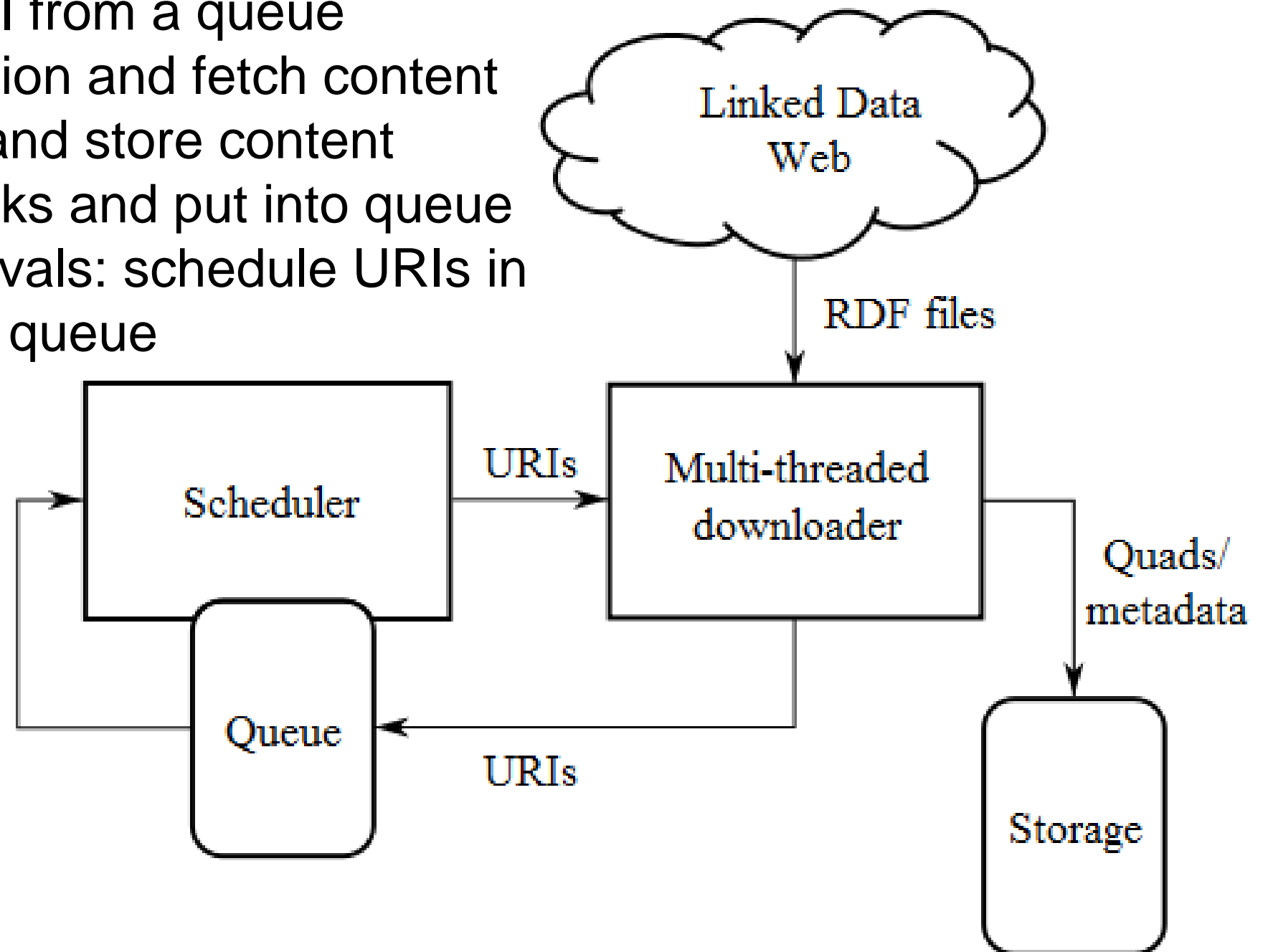
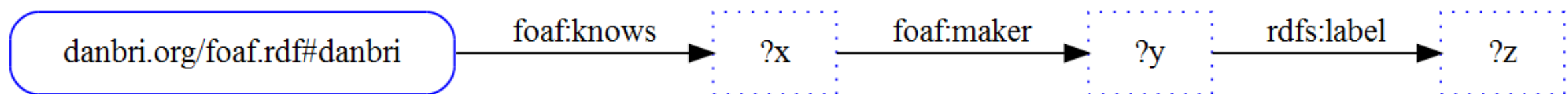


Image: Carlos Castillo: Effective Web Crawling, via Wikipedia

Conjunctive Queries / BGP

AIFB 

```
<http://danbri.org/foaf.rdf#danbri> foaf:knows ?x .
?x foaf:maker ?y .
?y rdfs:label ?z .
```



- Built on triple patterns containing variables $(?, ?, ?)$, $(s, ?, ?)$, $(?, p, ?)$, $(?, ?, o)$, $(s, p, ?)$, $(?, p, o)$, (s, p, o)
- Variables are bound during query evaluation
- Query evaluation results in a set of variable bindings

?x	?y	?z
http://www.w3.org/People/Berners-Lee/card#i	http://www.w3.org/DesignIssues/LinkedData.html	Linked Data - Design Issues - W3C

RDF Storage (1/4)

Horizontal table-per-class



Class: **Car**

plate	model	doors	owner	series	img	seeAlso
234HSJ	Toyota	5	Frank	Corolla	x.jpg	toyota.com
<div> <div> <div>Pros:</div> <ul style="list-style-type: none"> Fast for certain queries, esp. “star shaped” queries Little redundancy in cells </div> <div> <div>Cons:</div> <ul style="list-style-type: none"> Becomes very sparse for larger schema <ul style="list-style-type: none"> Lots of nulls needed Special handling needed for multi-valued attributes </div> </div>						
Frank	234HSJ	France	1.92		a.svg	...
Jill	148YUI	Ireland	1.75			...
Jim	923HIJ	US & UK		black	b.svg	

RDF Storage (2/4)

Vertical triple table

subject	predicate	object
Frank	ownsCar	234HSJ

- Pros:
 - No more nulls needed
 - Flexible for updates (even to data-schema)
 - Multi-valued attributes no problem
- Cons:
 - Lot's of self-joins
 - Lot's of redundancy in the cells

234HSJ	series	Corolla
--------	--------	---------

RDF Storage (3/4)

Vertical table per prop.

Property: **model**

subject	object
234HSJ	Toyota
923HIJ	Ford
242HFI	Fiat
541PJH	Porsche
148YUI	Jeep

Property: **ownsCar**

subject	object
Frank	234HSJ
Jill	148YUI
Jim	923HIJ
Joan	541PJH
Mary	242HFI

Property: **type**

subject	object
234HSJ	Car
923HIJ	Car
242HFI	Car
541PJH	Car
148YUI	Car

- Pros:
 - Less redundancy
- Cons:
 - Potentially many tables
 - New property = new table
 - Assumes predicate always known

RDF Storage (4/4)

Hybrid

AIFB

Class: **Car**

Property: **seeAlso**

plate	model	doors	owner	series
234HSJ	Toyota	5	Frank	Corolla
923HIJ	Ford	3	Jim	Ka
242HFI	Fiat	3	Mary	Tempura
541PJH	Pors			
148YUI	Jeep			

subject	object
234HSJ	toyota.com
923HIJ	ford.com
242HFI	fiat.com
	porsche.com
	jeep.com
	joan.com

- Pros:
- ~Depends
- Cons:
- Likely to be more costly to manage

name	ownsCar			
Frank	234HSJ	France	1.92	
Jill	148YUI	Ireland	1.75	
Jim	923HIJ	US & UK		black
Joan	541PJH	India	1.75	
Mary	242HFI	Scotland		

Property: **img**

subject	object
234HSJ	x.jpg
148YUI	i.png
Frank	a.svg
Jim	b.svg

Triple stores vs. Quad stores



1. Triple stores

- Only service simple RDF triple patterns
 - ♦ `?s rdf:type foaf:Person .`
 - ♦ `aidan ?p galway .`
 - ♦ `?s ?p ?o .`
 - ♦ ...

2. Quad stores/SPARQL engines

- Also service patterns involving named graphs
- Typical for indexing data from multiple sources
- Needed for SPARQL querying!!
 - ♦ `GRAPH ?g { ?s rdf:type foaf:Person }`
 - ♦ `GRAPH foaf.rdf { aidan ?p galway }`
 - ♦ `FROM graph1.rdf ... WHERE { ?s ?p ?o . }`

Building a full Quad index



- (subject, predicate, object, graph)
 - graph sometimes called context
- $2^4 = 16$ patterns to service!

No	Access pattern	No	Access pattern
1	(?:?:?:?)	9	(s?:o:c)
2	(s?:?:?)	10	(?:?:o:c)
3	(s:p?:?)	11	(?:?:o:?)
4	(s:p:o:?)	12	(?:?:?:c)
5	(s:p:o:c)	13	(s?:?:c)
6	(?:p?:?)	14	(s:p?:c)
7	(?:p:o:?)	15	(?:p?:c)
8	(?:p:o:c)	16	(s?:o:?)

Index Patterns

Six prefix-indexes for quads

- Requires six indexes to service all 16 quad patterns
 - assuming prefix lookups
 - ♦ e.g., can lookup SP in SPOC index... (~indexes should be sorted)

GRAPH ?c { :ted?p ?o }
 pattern: (s:?:?:?)
 index: spoc

GRAPH :dblp { :ted ?p ?o }
 pattern: (s:?:?:c)
 index: csp

GRAPH ?c { ?s rdf:type foaf:Person }
 pattern: (?p:o:?)
 index: poc

spoc	poc	ocs
(?:?:?:?)	(?:p:?:?)	(?:?:o:?)
(s:?:?:?)	(?:p:o:?)	(?:?:o:c)
(s:p:?:?)	(?:p:o:c)	(s:?:o:c)
(s:p:o:?)		
(s:p:o:c)		
csp	cp	os
(?:?:?:c)	(?:p:?:c)	(s:?:o:?)
(s:?:?:c)		
(s:p:?:c)		

Index Implementation

Blocked, Sorted File

AIFB 

- Sorted, compact
 - Combinatorial
 - Read optimized
 - Bulk-load
- Pros:**

 - !Fast! sequential reads (no OIDs)
 - Fast load
 - Good for simple queries, lots of results
 - Configurable block sizes

Cons:

 - Does not support updates!
 - Does not support updates!!!

```
...
http://deri.org/ dc:title
http://jfdk_id=47 foaf:
http://www.deri.org/ dc:title
...
```

```
http://jfdk_id=59 foaf:name "Dieter Fensel" http://deri.at/...foaf.rdf
...
```

```
http://www.deri.org/ dc:title "DERI" http://www.deri.ie/index.rdf
...
```

```
...
```

sparse index in memory

quadruples stored in blocks on disk, sorted according to natural ordering

Dictionary Encoding / Object IDs

Data Table

subject	predicate	object
1	2	3

Dictionary

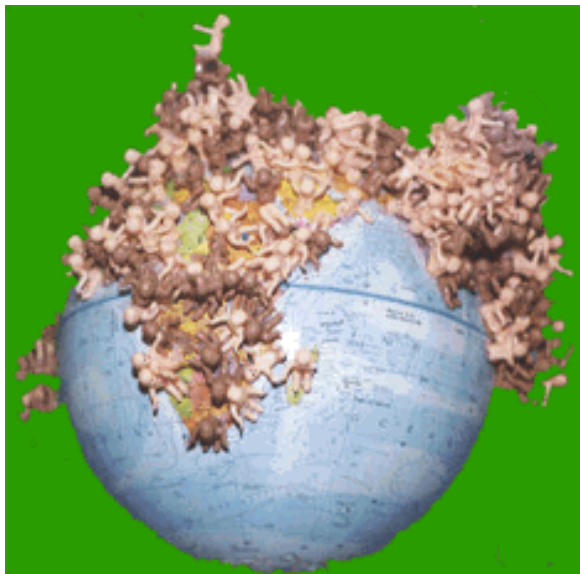
OID	String
1	Frank

1	
1	Pros:
1	Can load more data in memory
1	Faster to compute joins
1	Better RISC-style processing [RDF-3X]
1	Smaller on-disk footprint
1	Cons:
1	Maintain a potentially massive dictionary
1	Slower to externalise streaming results
1	FILTERs optimisations are trickier
...	

Query Processing (1/4)



```
?person foaf:name ?name .  
?person foaf:based_near dbpedia:Karlsruhe .  
:aidan foaf:knows ?person .
```



7 billion



0.3 million



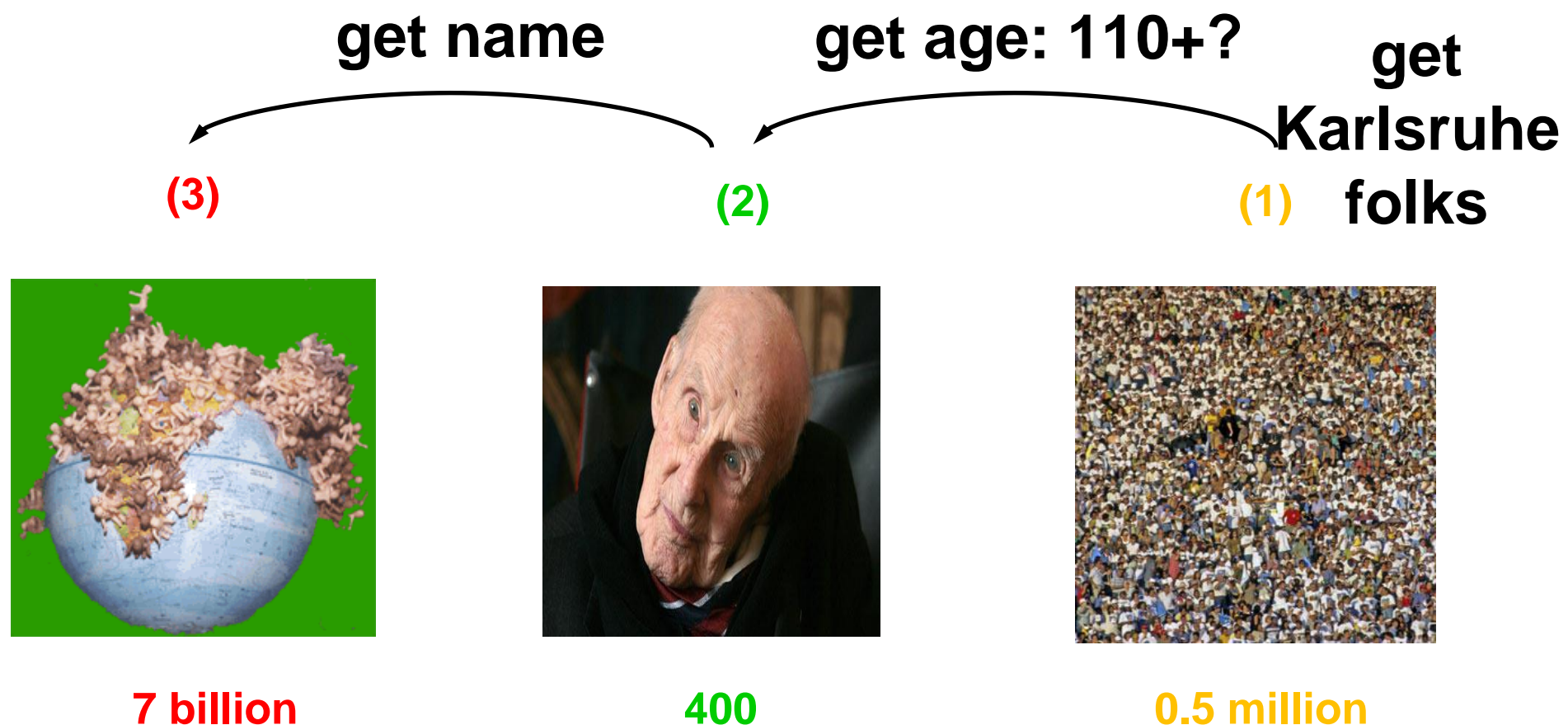
2

Query Processing (2/4)

Push FILTER expressions

```
?person foaf:name ?name .  
?person foaf:age ?age .  
?person foaf:based_near dbpedia:Karlsruhe .  
FILTER ( ?age > 110 )
```

- Filter intermediate results as soon as possible



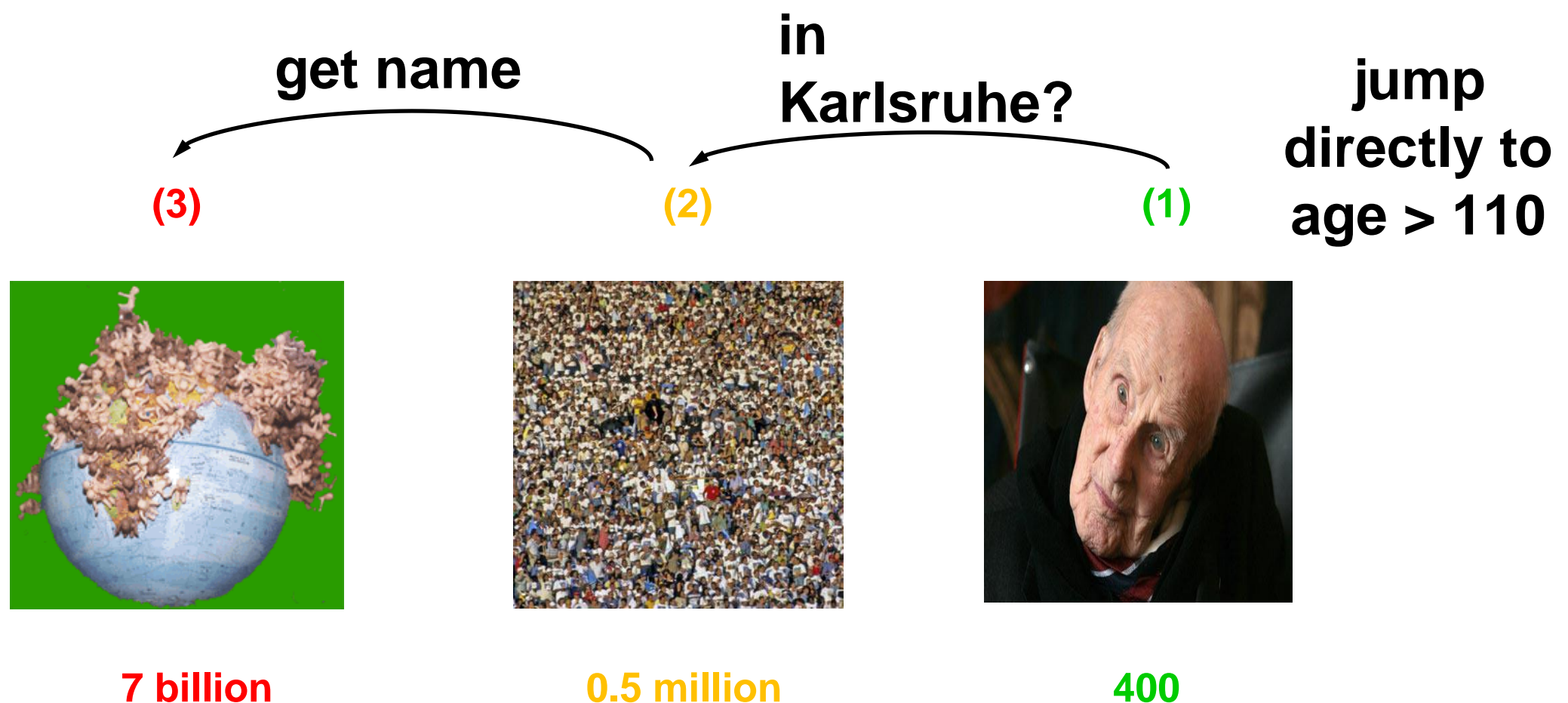
Query Processing (3/4)

Use Index, Push FILTER



```
?person foaf:name ?name .  
?person foaf:age ?age .  
?person foaf:based_near dbpedia:Karlsruhe .  
FILTER ( ?age > 110 )
```

- If you have the right index (sorted by PO or [maybe] OP prefix with numeric order on O), seek directly to 110+ for age



Query Processing (4/4)

No keyword search support in SPARQL

...give me resources mentioning “Bonn ISWC”

- Can use **FILTER/REGEX** in SPARQL...

```

                ?s ?p ?o , ?p2 ?o2 .
    FILTER ( REGEX(?o, "Bonn")  && REGEX(?o2, "ISWC") )
  
```

- *...but that sucks and is expensive*
- *...need an IR-style inverted index, not FILTER*
- Most SPARQL engines support custom keyword index/syntax
 - Against literals or concatenation of literals
 - Apache Lucene! ...perfect for the job!

Plan



XML und URIs
Einleitung in RDF
RDF Schema
Logik – Grundlagen
Semantik von RDF(S)
SPARQL – Syntax und Intuition
Semantik von SPARQL
Linked Data
Semantic Search
OWL – Syntax und Intuition I
OWL – Syntax und Intuition II
OWL – Semantik und Reasoning
Konjunktive Anfragen und Regelsprachen
Applications

Attribution



- Slides erstellt von Andreas Harth, Aidan Hogan, Spyros Kotoulas, Jacopo Urbani für das Tutorial “**Scalable Integration and Processing of Linked Data**”
 - siehe <http://sild.cs.vu.nl/>